Journal of Transport & Health 35 (2024) 101778

Contents lists available at ScienceDirect

Journal of Transport & Health

journal homepage: www.elsevier.com/locate/jth

Validating the Vitals assessment: A replication study on cognitive assessments and commercial driving risk

Alice Elizabeth Atkin [a,b,*], Daniel Scott [c,**], Anthony Singhal [a,b,***]

[a] Neuroscience and Mental Health Institute, University of Alberta, Canada
[b] Department of Psychology, University of Alberta, Canada
[c] Impirica, Inc, Canada

1. Introduction

In a recently published article, the authors report a logistic regression model for predicting on-road driving risk in commercial drivers depending upon their performance on a battery of four cognitive and sensorimotor tasks (Scott et al., 2023). This tablet-based task battery, known as Vitals, consists of four tasks, which assess reaction time, decision making, working memory, and motor control, while on-road driving risk was assessed by modifying a standardized evaluation that is designed to identify errors characteristic of cognitive impairment risk. Past research has shown that this methodology offers good predictive value in assessing driver risk across a range of populations and vehicle types (Bakhtiari et al., 2020; Dobbs, 2013; Choi et al., 2015; Korner-Bitensky and Sofer, 2009). In Scott et al. (2023), we trained our regression model on a dataset of 1343 drivers recruited from a variety of commercial, judicial, medical, and academic sources, before validating it on our study sample, which consisted of 2167 commercial drivers recruited through their employers. Although we did not train our model exclusively on commercial drivers, it nevertheless showed good predictive value for our validation sample, with a strong relationship observed between a driver's Vitals score and their likelihood of on-road evaluation failure (Fig. 1). This result suggests that the Vitals could be an effective screening tool for risky on-road driving in commercially licensed drivers.

The study reported in Scott et al. (2023) was limited to a sample size of 2167 commercial drivers, but data collection continued following the cut-off date for inclusion in that study, and we now have a considerably larger sample of data to analyze. In part, we can affirm the credibility of a scientific result by the ability to replicate that result with new data collected under similar circumstances (Cesario, 2014; Nosek and Errington, 2020). Replication is not the only or essential method for affirming the credibility of a result, and even a replicable result can lack credibility if the methods lack validity (Devezer et al., 2019). Nevertheless, our ability to continuously collect data for this project offers a worthwhile opportunity to directly replicate the results of the model presented in Scott et al. (2023). Furthermore, the proprietary nature of that study's tasks and dataset confers a responsibility on the authors to affirm or disaffirm the credibility of our study, given that direct replication cannot be performed by independent researchers.

Using our new, post-validation sample, we re-ran the same logistic regression model reported in Scott et al. (2023) and then compared the model's measures of positive predictive value (PPV), negative predictive value (NPV), false negative rate (FN%), and false positive rate (FP%) against those from our initial validation sample. Similar to the concept of a noninferiority trial in medicine (Mauri and D'Agostino, 2017), our aim with this new analysis is not to show that our model performs better on one sample versus the other, but instead to show that the model performs just as well on either sample. By performing this analysis, we hope to reinforce the

* Corresponding author: Neuroscience and Mental Health Institute, University of Alberta, Canada.
** Corresponding author.
*** Corresponding author: Neuroscience and Mental Health Institute, University of Alberta, Canada.
E-mail addresses: aliceatkin@ualberta.ca (A.E. Atkin), d.scott@impirica.tech (D. Scott), asinghal@ualberta.ca (A. Singhal).

https://doi.org/10.1016/j.jth.2024.101778
Received 24 April 2023; Received in revised form 25 January 2024; Accepted 12 February 2024
Available online 6 March 2024

credibility and utility of our model by validating it a second time on an independent sample of commercial drivers.

2. Methods

For the post-validation sample, we recruited 1924 commercial truck, heavy, and light vehicle drivers from various organizations across Canada. All participants gave informed consent to have their data collected for this study. Completion of the Vitals assessment and on-road assessment was conducted as part of each participant's job duties or as a condition of their hiring by an organization, but participants could refuse consent for their data to be included in this study. The study was approved by the University of Alberta Health Research Ethics Board and was performed in compliance with relevant laws and institutional guidelines, as well as the Declaration of Helsinki.

Full details about the study methodology and procedure can be found in Scott et al. (2023), including details about the procedure used to remove age affects from the logistic model. In brief, each participant completed the Vitals cognitive risk assessment tool as well as an on-road evaluation using their vehicle. The Vitals tool consists of four tasks which measure reaction time, decision-making, working memory, and sensorimotor control. The Vitals is performed on a tablet and takes approximately 30 min to complete. After completing the Vitals, each participant was assigned a Vitals Score representing our model's predicted likelihood that they would fail their on-road evaluation (Fig. 1). The on-road evaluation was conducted by a trained evaluator according to a standardized assessment tool known as the Commercial On-Road Evaluation (CORE). The CORE is a modified version of the DriveABLE On-Road Evaluation (DORE), with different categories of assessment depending upon vehicle type. The CORE uses the same assessment and course design methodology as the DriveABLE On-Road Evaluation (DORE). In the original development of the DORE, errors were categorized as either typical or hazardous (Dobbs, 1997, 2013; Dobbs et al., 1998). In the DORE research, the accumulation of typical errors did not explain differences between healthy and cognitively impaired drivers. By contrast, the accumulation of hazardous errors did explain such differences, and were characteristic of cognitively impaired drivers. In this study, we inherit the cutpoints for hazardous error failure from the DORE. Using the logistic regression with age effects removed, these driving outcomes were effectively predicted using the Vitals assessment. The CORE takes approximately 1 h to complete, and each participant was classified as "pass", "borderline", or "fail", depending upon the number of hazardous errors.

After recruiting participants for the post-validation sample (n = 1924), we observed comparable proportions of each vehicle type, age range, and outcomes on the CORE relative to the validation sample (n = 2167) reported in Scott et al. (2023), as seen in Tables 1 and 2.

3. Results

To analyze our two samples, we used a two-sided z-test of proportions to evaluate differences between the Vitals' predictive value for the validation and post-validation samples. The set of accuracy metrics used in the original validation paper is included: positive predictive value (PPV), negative predictive value (NPV), false negative rate (FN%) and false positive rate (FP%) (Table 3). There were no significant differences found in predictive value or error rate between the validation and post-validation samples (Table 4 and Fig. 1).

4. Conclusion

In this study, we directly replicated the conclusions of Scott et al. (2023) using a new sample of data collected from the same organizations. The same methods, procedure, and model were followed for this post-validation study as in the original validation

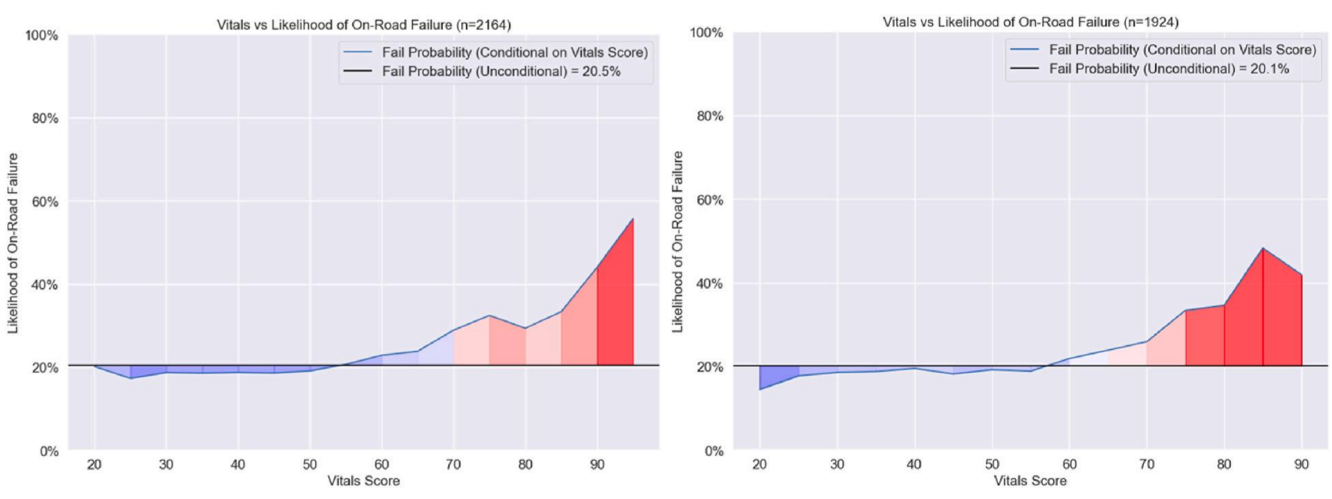

Fig. 1. Consistent with the validation study (left, n = 2164), the post-validation study demonstrated an increase in real-world driving risk for drivers who had a high Vitals assessment score (right, n = 1924). Adapted from Scott et al. (2023).

Table 1
On-road pass/borderline/fail sample sizes for drivers of heavy vehicles, light vehicles, and trucks in the validation dataset (n = 2167). Reprinted from Scott et al. (2023).

	Driving Pass	Driving Borderline	Driving Fail	N	Age (M ± SD)
Heavy	273	65	77	415	41.7 ± 13.0
Light	919	346	327	1592	36.95 ± 10.7
Trucking	99	22	39	160	43.69 ± 12.57

Table 2
On-road pass/borderline/fail sample sizes for drivers of heavy vehicles, light vehicles, and trucks in the post-validation dataset (n = 1924).

	Driving Pass	Driving Borderline	Driving Fail	N	Age (M ± SD)
Heavy	212	53	57	322	40.5 ± 13.1
Light	644	218	284	1146	40.0 ± 11.4
Trucking	340	70	46	456	37.4 ± 13.1

Table 3
Proportions for each of the accuracy metrics in the validation and post-validation study groups. None of the accuracy metrics in the post-validation study were worse than the validation study. Adapted from Scott et al. (2023).

Group	PPV	NPV	FN%	FP%
Validation Study	0.406	0.819	0.171	0.034
Post-Validation Study	0.380	0.816	0.175	0.031

Table 4
Two-sided Z-test of proportions (z statistic, p) for each of the accuracy metrics in the validation and post validation groups. None of the accuracy metrics in the post-validation study were statistically different than the validation study.

Statistic	PPV	NPV	FN%	FP%
z	−1.55	−0.17	−0.28	0.57
p-value	0.1215	0.8636	0.7799	0.5664

study. Participants in each sample completed the Vitals cognitive risk assessment tool and an on-road driving evaluation, and the likelihood of evaluation failure due to hazardous error was predicted on the basis of Vitals performance. The model's PPV, NPV, FN%, and FP% were measured for each sample (Table 3), and no differences were observed between the two samples on any measure (Table 4). In both samples, drivers who were predicted to have the highest likelihood of failing their on-road evaluation were in fact most likely to fail (Fig. 1). Approximately 20.4% of drivers in our sample failed the on-road evaluation, and were therefore classified as being at risk of unsafe driving. Such a percentage of at-risk drivers is consistent with insurance industry assessments of driver risk. For instance, according to insurance industry data reported by Sapna Isotupa et al. (2019) , at least 23.4% of licensed drivers in the authors' home province of Alberta can be classified as having elevated risk and charged increased rates, depending on the classification system used by the insurer. While these systems classify risk based on a different variable than that used by our study (recent claims history versus performance on the CORE) and encompasses all drivers, not just commercial drivers, our study results do not classify a significantly larger percentage of drivers as being at-risk than might reasonably be expected, given this industry data. The results of this analysis therefore affirm the conclusions of Scott et al. (2023) that poor performance on the Vitals is a reliable basis for predicting on-road risk in commercial drivers. These results, obtained under common model parameters, imply that a generalized predictive power exists for all three vehicle groups (heavy, light, and trucking), mitigating any overfitting risks associated with separate modeling for each group.

This study adds to the growing evidence that off-road cognitive risk assessments can be reliably used to predict risky driving across a range of populations and vehicle types (Bakhtiari et al., 2020; Dobbs, 2013; Choi et al., 2015; Korner-Bitensky and Sofer, 2009; Scott et al., 2023). Given that the Vitals tool is currently being implemented by transportation organizations as part of their hiring procedures and fitness-for-duty assessments, it is important to analyze the value of this tool for predicting driver risk within those same settings and populations (Scott et al., 2023). By replicating our model's results using a new sample of drivers, this study provides confirmatory evidence of the Vitals' utility for predicting on-road evaluation failure in commercial drivers. Overall, our results suggest that use of the Vitals can provide relevant and reliable information about on-road driving risk in commercial drivers when implemented as part of an overall hiring or fitness assessment. In conducting this post-validation analysis of our published model using a new sample, we hope to join other researchers in promoting more rigorous and thorough methods of affirming the credibility of one's experimental results, and to encourage other researchers to directly replicate their own studies when feasible.

Role of the funding source

Funding for the conduct of this study and preparation of the article, including data collection, analysis, and writing, was provided by a Natural Sciences and Engineering Research Council (NSERC) Collaborative Research and Development grant awarded to Dr. Anthony Singhal and Impirica, and was also provided by Impirica.

CRediT authorship contribution statement

Alice Elizabeth Atkin: Conceptualization, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Daniel Scott:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Anthony Singhal:** Conceptualization, Funding acquisition, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Anthony Singhal reports financial support was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC). Daniel Scott reports financial support, administrative support, and equipment and supplies were provided by Impirica. Daniel Scott reports a relationship with Impirica that includes: employment and equity or stocks.

Data availability

The data that has been used is confidential.

References

Bakhtiari, R., Tomczak, M.V., Langor, S., Scanlon, J.E.M., Granley, A., Singhal, A., 2020. Application of tablet-based cognitive tasks to predict unsafe drivers in older adults. TRIP 4, 100105. https://doi.org/10.1016/j.trip.2020.100105.

Cesario, J., 2014. Priming, replication, and the hardest science. Perspect. Psychol. Sci. 9 (1), 40–48. https://doi.org/10.1177/1745691613513470.

Choi, S.Y., Yoo, D.H., Lee, J.S., 2015. Usefulness of the driveABLE cognitive assessment in predicting the driving risk factor of stroke patients. J. Phys. Ther. Sci. 27 (10), 3133–3135. https://doi.org/10.1589/jpts.27.3133.

Devezer, B., Navarro, D.J., Vandekerckhove, J., Buzbas, E.O., 2019. The Case for Formal Methodology in Scientific Reform, vol. 8. Royal Society Open Science, 200805. https://doi.org/10.1098/rsos.200805.

Dobbs, A.R., 1997. Evaluating the driving competence of dementia patients. Alzheimer Dis. Assoc. Disord. 11 (1), 8–12.

Dobbs, A.R., 2013. Accuracy of the DriveABLE cognitive assessment to determine cognitive fitness to drive. Can. Fam. Physician 59 (3), e156–e161.

Dobbs, A.R., Heller, R.B., Schopflocher, D., 1998. A comparative approach to identify unsafe older drivers. Accid. Anal. Prev. 30 (3), 363–370. https://doi.org/10.1016/S0001-4575(97)00110-3.

Korner-Bitensky, N., Sofer, S., 2009. The DriveABLE Competence Screen as a predictor of on-road driving in a clinical sample. Aust. Occup. Ther. J. 56 (3), 200–205. https://doi.org/10.1111/j.1440-1630.2008.00749.x.

Mauri, L., D'Agostino, R.B., 2017. Challenges in the design and interpretation of noninferiority trials. N. Engl. J. Med. 377, 1357–1367. https://doi.org/10.1056/NEJMra1510063.

Nosek, B.A., Errington, T.M., 2020. What is replication? PLoS Biol. 18 (3), e3000691 https://doi.org/10.1371/journal.pbio.3000691.

Sapna Isotupa, K.P., Kelly, M., Kleffner, A., 2019. North Am. Actuar. J. 23 (3), 395–411. https://doi.org/10.1080/10920277.2019.1572524.

Scott, D., Atkin, A.E., Granley, A., Singhal, A., 2023. The utility of cognitive testing to predict real world commercial driving risk. TRIP 18, 100783. https://doi.org/10.1016/j.trip.2023.100783.